



**[biblio.ugent.be](http://biblio.ugent.be)**

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Multi-loop Quality Scalability based on High Efficiency Video Coding

Glenn Van Wallendael, Jan De Cock, Marta Mrak, Rik Van de Walle

Proceedings of Picture Coding Symposium, Krakow, Poland, May 2012.

**To refer to or to cite this work, please use the citation to the published version:**

**G. Van Wallendael, J. De Cock, M. Mrak, R. Van de Walle, "Multi-loop Quality Scalability based on High Efficiency Video Coding", Proceedings of Picture Coding Symposium, Krakow, Poland, May 2012**

# Multi-loop Quality Scalability based on High Efficiency Video Coding

Glenn Van Wallendael, Jan De Cock, Rik Van de Walle  
Multimedia Lab - IBBT  
Department of Electronics and Information Systems  
Ghent University, Ghent, Belgium  
e-mail: glenn.vanwallendael@ugent.be

Marta Mrak  
British Broadcasting Corporation  
Research and Development Department  
London, United Kingdom  
e-mail: marta.mrak@bbc.co.uk

**Abstract**—Scalable video coding performance largely depends on the underlying single layer coding efficiency. In this paper, the quality scalability capabilities are evaluated on a base of the new High Efficiency Video Coding (HEVC) standard under development. To enable the evaluation, a multi-loop codec has been designed using HEVC. Adaptive inter-layer prediction is realized by including the lower layer in the reference list of the enhancement layer. As a result, adaptive scalability on frame level and on prediction unit level is accomplished. Compared to single layer coding, 19.4% Bjøntegaard Delta bitrate increase is measured over approximately a 30dB to 40dB PSNR range. When compared to simulcast, 20.6% bitrate reduction can be achieved. Under equivalent conditions, the presented technique achieves 43.8% bitrate reduction over Coarse Grain Scalability of the SVC - H.264/AVC-based standard.

**Index Terms**—quality scalability, High Efficiency Video Coding (HEVC), multi-loop compression

## I. INTRODUCTION

High Efficiency Video Coding (HEVC) is the new video coding standard developed by a joint effort between ISO/IEC and ITU-T and succeeding H.264/AVC. When objectively comparing the current HEVC compression performance with H.264/AVC High Profile [1], already 44% bitrate reduction [2] is obtained. Subjective evaluation is likely to show even better performance. Given the advanced techniques and high compression ratios obtained by HEVC (i.e., a single layer coding scheme), an intriguing question is how much the performance of a scalable variant would be affected?

With a scalable extension like Scalable Video Coding (SVC) [3], it is possible to adapt the video stream to a different resolution, frame rate, or quality. Consequently, the video stream can be modified to support varying network conditions, end-user devices, and applications. Especially when coping with changing network capacity, quality scalability offers significant flexibility increasing end user's Quality of Experience (QoE) [4]. Because scalable techniques largely depend on the underlying single layer compression, with the upcoming HEVC standard, new possibilities for scalable improvement should be investigated. While intensive research is still being carried out on single layer coding, some preliminary investigations on a scalable extension of HEVC are taking place. Several contributions have already been made to the Joint Collaborative Team on Video Coding (JCT-VC) investigating

spatial scalability performance of HEVC [5] [6]. In this paper, a simple yet effective quality scalable design based on HEVC is investigated.

A general introduction on relevant HEVC coding concepts is given in Section II. Section III summarizes proposed changes to HEVC that are introduced for quality support and evaluation. Finally, results will be given in Section IV followed by concluding remarks in Section V.

## II. HIGH EFFICIENCY VIDEO CODING

Similar to H.264/AVC [1], HEVC is a hybrid block based video coder. The coding structure as it will be described here corresponds to the fourth Working Draft (WD) [7] of HEVC.

As in H.264/AVC, decoding-related parameters are available in several groups. Sequence Parameter Sets (SPS) and Picture Parameter Sets (PPS) carry the information on a sequence and picture levels. Further information needed for decoding the actual video information can be found in the slice header of every slice.

Concerning the actual video content, the picture is subdivided into Large Coding Units (LCU), which are typically 64x64. However, the LCU size can be changed in the SPS. Corresponding to the macroblock scanning in H.264/AVC, LCUs are encoded in a raster scan order. Each LCU is subdivided into smaller Coding Units (CU) according to a quadtree structure [8] [9]. This subdivision can recursively continue until the smallest CU size of 8x8 is obtained. Information contained by these CUs is transmitted in a Z-scan order. On the CU level, the corresponding prediction type is defined as well as a flag indicating whether the block coding is skipped or not. Analogous to macroblocks in H.264/AVC, CUs can be inter or intra predicted. Only inter prediction will be considered further in this paper since, in the presented approach, intra prediction techniques are left unchanged compared to HEVC.

Every leaf CU of the quadtree contains one or more Prediction Units (PU) and Transform Units (TU) [10]. In TUs, transform related information and residual data is signalled. On the TU level, a restricted form of quadtree splitting can be applied corresponding to rate-distortion optimality [11].

Prediction structure defines CU split into square or non-square blocks labeled as 2Nx2N, 2NxN, Nx2N, NxN, or asymmetric. In HEVC terminology, a 2Nx2N size means the

entire CU is one prediction unit.  $N \times N$ , on the other hand, indicates that the CU is split in four quarter sized PUs. As a result,  $2N \times N$  and  $N \times 2N$  sizes represent rectangular sized PUs.

Motion information is signalled at the PU level. Every PU is predicted from one or two reference pictures by means of motion compensated prediction. These reference pictures are enlisted in two reference picture lists (L0 and L1) and a Combined List (LC). The reference picture lists are filled with available pictures in a default way. Changing this default behavior can be facilitated with reference picture list modification signalling, which communicates modified lists to the decoder.

The motion information signalled in the PU consists of a reference index, a motion vector prediction index, and a motion vector difference. First potential motion vector predictors are enlisted by the decoder. The motion vector prediction index points to the most efficient predictor in the created list. This motion vector predictor is further adjusted with the signalled motion vector difference to obtain the absolute motion vector. For the calculation of motion vector predictors, neighboring motion vectors are considered. Before being added to the candidate list, these motion vectors are scaled corresponding to the gap between the current picture and the referenced picture. For example, when a neighboring motion vector points to a picture that is twice as far compared to the distance with the currently used reference frame, then this motion vector is halved.

Additional to possibilities for partitioning, a vast amount of extra tools is included compared to H.264/AVC. In this way, HEVC is able to get its superior performance compared to H.264/AVC.

### III. ENABLING QUALITY SCALABILITY BASED ON HEVC

First, an explanation will be given about the difference between a single-loop scalable codec like SVC and a multi-loop scalable codec as proposed. Then, in order to enable quality scalability based on HEVC, changes in picture coding order will be explained. Finally, techniques to enable inter-layer prediction will be elaborated on.

With multi-loop scalable compression, the lower layer is decoded entirely before the Enhancement Layer (EL) can be reconstructed. The main advantage of this approach is that decoded pixel information from the lower layer can be used during the inter-layer prediction process of the enhancement layer. On the other hand, with single loop decoding, no motion compensation is done in lower layers of the scalable video stream. Consequently, decoded pixels from inter-predicted parts of the lower layer are not available for prediction of higher layers. Without access to inter-predicted pixel information, it is still possible to improve the compression of higher layers with residual signal and motion information. Single-loop decoding is in general less complex for the decoder, but at the cost of compression efficiency.

To obtain an efficient multi-loop scalable implementation of HEVC, the HEVC picture decoding order was modified to

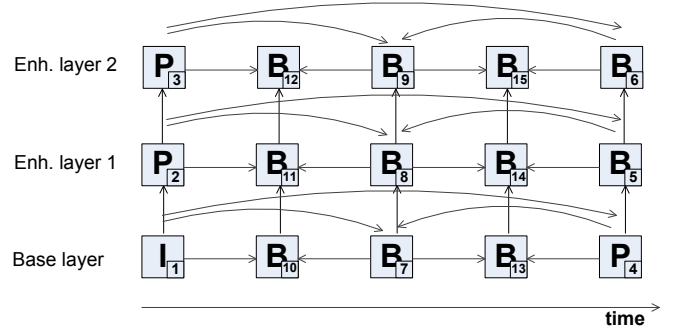


Fig. 1. Scalable frame type and coding order.

first process every quality layer of the same picture before advancing to following pictures in decoding order. Consequently, Base Layer (BL) pixel information can be made accessible for the higher layer decoding process. The coding order resulting from this decision is illustrated with the numbering in Fig. 1.

Inter-layer prediction is enabled entirely by means of motion compensation. As a result, all flexibility available in the HEVC specification can be used to adaptively enable inter-layer prediction. On two levels of granularity, inter-layer prediction can be controlled.

First of all, by means of reference picture list modification signalling, inter-layer prediction can be enabled or disabled on a picture level. Enabling inter-layer prediction can be done by signalling a reference picture list modification message telling the decoder to include the lower layer in the reference list. In the absence of such a message, no inter-layer prediction will be used. This flexibility is for instance beneficial when decoding a hierarchically encoded video stream with large quality difference between layers. In this scenario, inter-layer prediction can be disabled for example at the highest temporal layer. At these positions, pictures are closest to already decoded pictures and therefore, temporal prediction is preferred to inter-layer prediction. Disabling inter-layer prediction for these pictures leaves room for extra temporal predictions in the reference lists, optimizing compression efficiency of the video stream.

When inter-layer prediction is enabled at the frame level by including the lower layer in the reference list, additionally, inter-layer prediction can be adaptively signalled on the PU level. By using the reference index pointing to this layer, inter-layer prediction will be used instead of temporal prediction. In the encoder, this choice is made dependent on rate-distortion performance.

For insertion of the lower layer in the reference list of the currently processed picture, three different scenarios can occur depending on the prediction type of the BL, namely I, P, or B predicted BL. This is also illustrated with the prediction types indicated in the pictures in Fig.1. When there is an I picture in the BL, the  $N^{th}$  EL can only use layer  $N - 1$  for prediction. As a result, in this scenario, EL pictures will be unidirectional Predicted (P) pictures. Consequently, each reference picture list at level  $N$  contains the picture from layer

TABLE I  
JCT-VC TEST SEQUENCES

Class	Sequence name	Frame count	Frame rate	Resolution
B	Kimono	240	24fps	1920x1080
B	ParkScene	240	24fps	1920x1080
B	Cactus	500	50fps	1920x1080
B	BQTerrace	600	60fps	1920x1080
B	BasketballDrive	500	50fps	1920x1080
C	BasketballDrill	500	50fps	832x480
C	BQMall	600	60fps	832x480
C	PartyScene	500	50fps	832x480
C	RaceHorses	300	30fps	832x480
D	BasketballPass	500	50fps	416x240
D	BQSquare	600	60fps	416x240
D	BlowingBubbles	500	50fps	416x240
D	RaceHorses	300	30fps	416x240
F	BasketballDrillText	500	50fps	832x480
F	ChinaSpeed	500	30fps	1280x720
F	SlideEditing	300	30fps	1280x720
F	SlideShow	500	20fps	1280x720

$N - 1$  as reference. Layers enhancing a P predicted BL will become B predicted pictures, because the lower layer can be added to a reference list. In the presented approach, layer  $N - 1$  is added as the first picture in the L0 reference picture list. For B predicted BL pictures, a similar procedure is followed, but it must be considered that the reference lists can already be filled up. In L0, the lower layer is included as an additional reference at the cost of the last reference picture in this list.

As already explained in the previous section, in HEVC, a motion vector predictor is calculated after weighting the distance between the pictures in display order. The weighting function is changed to assume a zero motion vector when inter-layer prediction is used. This results from the assumption that, during inter-layer prediction, the most resembling pixel information can be found on the same spatial location in the lower layer.

#### IV. RESULTS

Our evaluation conditions are derived from the common conditions as applied during the HEVC standardization process [12]. From these conditions, the restrictions from the random access high efficiency configuration are applied. This configuration implies that random access to the video stream is possible approximately every second. It also means that CABAC, hierarchical B prediction, Adaptive Loop Filtering (ALF), and 10-bit internal video representation are used. These common conditions also specify the Quantization Parameters (QP) (22, 27, 32, 37) on which the tests should be run. These QPs are assigned to the EL. Two layers are coded with a QP difference of five. A difference of five is chosen because then the QP gap between the BL and the EL is equal to the next considered test point. As a result, a realistic extrapolation for more than two layers can be made. The QP difference of five also challenges the scalable codec, because a big quality gap between layers decreases inter-layer prediction performance. On the other hand, a bigger QP difference would not be relevant, because for quality scalability, smaller quality differences are advised [13]. For higher quality differences a combination

TABLE II  
BD RATE PERFORMANCE OF SVC COMPARED TO SINGLE LAYER CODING AND SIMULCAST ENCODING.

Class	single layer	simulcast
B	24.4%	-15.5%
C	19.2%	-21.6%
D	18.7%	-22.0%
F	18.4%	-26.9%
Avg	20.4%	-21.1%

of spatial and quality scalability must be considered, but this is outside the scope of this paper.

The described scalable modifications are based on the HEVC Test Model (HM) version 4.0 [14]. Because of the quality difference between both layers, performance measurements showed suboptimal performance from including inter-layer prediction in all temporal layers. Consequently, the flexibility of reference picture list modification is utilized to disabled inter-layer prediction on the highest temporal layer.

To compare the performance of the proposed technique with SVC, simulation results from SVC, using JSVM reference software version 9.19.14 [15], under similar conditions are presented first. Identical conditions cannot be simulated, because some aspects of HEVC compression were not present in SVC. First of all, SVC is a single loop decoding technology in contrast to multi-loop decoding used in the proposed technique. There is also a large difference of coding tools and complexity between H.264/AVC based compression and HEVC based compression. For example, in SVC, Instantaneous Decoder Refresh (IDR) pictures are used to enable random access. In HEVC, better compression performance can be obtained with Clean Random Access (CRA). This is a random access technique based on open GOP encoding of I frames. In this paper, we have chosen to leave HEVC in its optimal configuration creating a more challenging configuration for the scalability tools. Results from comparing SVC with HEVC can still be regarded valid because under the conditions used, similar single layer gains are reported in a more fair comparison [2].

Under the conditions, both the performance of SVC and the proposed scalable modification of HEVC are evaluated as described next. For both codec generations, first a comparison is made with single layer coding. The single layer version has the same QP as the EL of the scalable variation. Therefore, this comparison represents the cost of introducing the flexibility of scalability to the video stream. As a second test, a comparison is made with simulcast encoding. With the simulcast configuration, two single layer versions are encoded at a quality of one of the layers of the scalable configuration. This result represents the gain of scalable coding when the flexibility of having different bitrates is required.

All the defined tests are run on the test sequences as listed in Table I. As a metric for evaluating the results, Bjøntegaard Delta (BD) [16] measurements are used. This measure gives an average bandwidth change over the evaluated QP range.

To make a comparison between the scalability performance of SVC and the proposed technique, in Table II, BD bitrate

TABLE III  
PERFORMANCE OF QUALITY SCALABLE HEVC COMPARED TO SINGLE  
LAYER CODING, SIMULCAST ENCODING, AND SVC.

Class	Sequence	single layer	simulcast	SVC
B	Kimono	16.7%	-21.8%	-62.2%
B	ParkScene	18.1%	-18.8%	-43.3%
B	Cactus	20.9%	-18.5%	-49.5%
B	BasketballDrive	18.2%	-20.6%	-58.5%
B	BQTerrace	13.6%	-17.3%	-44.0%
Average B		17.5%	-19.4%	-51.5%
C	BasketballDrill	22.0%	-19.3%	-42.7%
C	BQMall	22.5%	-19.4%	-59.7%
C	PartyScene	20.3%	-18.6%	-49.0%
C	RaceHorses	17.0%	-20.3%	-49.0%
Average C		20.5%	-19.4%	-50.1%
D	BasketballPass	20.2%	-20.4%	-12.8%
D	BQSquare	21.4%	-18.9%	-43.8%
D	BlowingBubbles	22.5%	-16.9%	-23.0%
D	RaceHorses	20.3%	-19.4%	-35.7%
Average D		21.1%	-18.9%	-28.8%
F	BasketballDrillText	23.3%	-19.1%	-36.2%
F	ChinaSpeed	19.9%	-20.1%	-33.4%
F	SlideEditing	9.9%	-36.4%	-35.9%
F	SlideShow	23.4%	-24.7%	-65.4%
Average F		19.1%	-25.1%	-42.7%
Average		19.4%	-20.6%	-43.8%

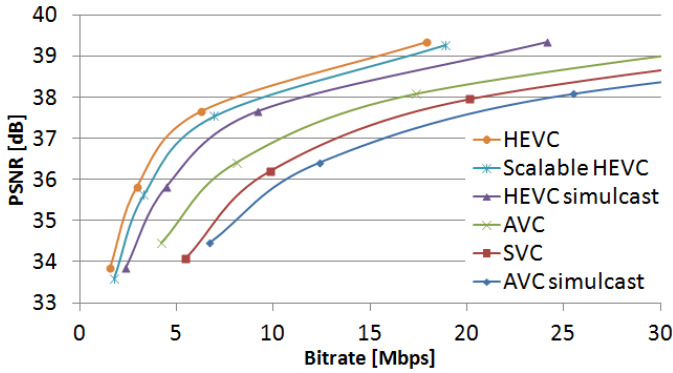


Fig. 2. Rate distortion curves of BasketballDrive sequence (1080p 50fps).

results of SVC compared to single layer H.264/AVC and compared to simulcast are shown. In this table, it can be observed that the cost of encoding a video stream as two scalable SVC layers adds a BD rate cost of 20.4%. The table also shows that SVC results in a 21.1% gain compared to a simulcast H.264/AVC configuration.

Equivalently, the BD rate performance of the proposed system compared to HEVC is shown in Table III. In the table, scalable coding results compared to single layer or simulcast can be observed. On average, the BD rate cost of having a scalable layer is 19.4% and the gain compared to simulcast HEVC is 20.6%. It can also be observed that BD bitrate savings are consistent over all tested sequences. In the last column of Table III, a comparison with SVC is made. On average, 43.8% BD bitrate saving is realized by using the proposed method. Finally, rate distortion performance of the BasketballDrive sequence is illustrated in Fig. 2. In this graph, the performance difference between H.264/AVC and HEVC technology is apparent.

## V. CONCLUSION

In this paper, a multi-loop quality scalable video codec, built on top of HEVC, is proposed. Quality scalability is obtained by including the decoded lower layer in the reference picture list of the enhancement layer. Consequently, by means of reference frame indexing, inter-layer prediction can be enabled or disabled on prediction unit level. With this technique, quality scalable coding in HEVC comes at a BD bitrate cost of 19.4% compared to single layer coding. Compared to simulcast, a BD bitrate decrease of 20.6% can be gained with the scalable solution. These results show that even with a more efficient single layer codec compared to H.264/AVC, still an equally BD rate efficient scalable extension can be obtained. Finally, when comparing the proposed technique with SVC, 43.8% BD bitrate reduction can be achieved.

## REFERENCES

- [1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [2] B. Li, G. J. Sullivan, and J. Xu, "Comparison of Compression Performance of HEVC Working Draft 4 with AVC High Profile," *JCTVC-G399*, Geneva, Switzerland, Nov. 2011.
- [3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [4] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized Rate-Distortion Extraction with Quality Layers in the Scalable Extension of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1186–1193, Sep. 2007.
- [5] D. Hong, W. Jang, J. Boyce, and A. Abbas, "Scalability Support in HEVC," *JCTVC-F290*, Torino, Italy, July. 2011.
- [6] H. Choi, J. Nam, and D. Sim, "Scalable structures and inter-layer predictions for HEVC scalable extension," *JCTVC-F096*, Torino, Italy, July. 2011.
- [7] B. Bross, W.-J. Han, J.-R. Ohm, G. Sullivan, and T. Wiegand, "WD4: Working Draft 4 of High-Efficiency Video Coding," *JCTVC-F803*, Torino, Italy, July. 2011.
- [8] W.-J. Han, J. Min, I.-K. Kim et al., "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1709–20, Dec. 2010.
- [9] M. Karczewicz, P. Chen, R. Joshi et al., "A Hybrid Video Coder Based on Extended Macroblock Sizes, Improved Interpolation, and Flexible Motion Representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1698–1708, Dec. 2010.
- [10] F. Bossen, V. Drugeon, E. Francois et al., "Video Coding Using a Simplified Block Structure and Advanced Coding Techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1667–1675, Dec. 2010.
- [11] D. Marpe, H. Schwarz, S. Bosse et al., "Video Compression Using Nested Quadtree Structures, Leaf Merging, and Improved Techniques for Motion Representation and Entropy Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1676–1687, Dec. 2010.
- [12] F. Bossen, "Common test conditions and software reference configurations," *JCTVC-F900*, Torino, Italy, July. 2011.
- [13] ISO/IEC JTC1/SC29/WG11, "Draft Requirements and Discussion on the scalable enhancement of HEVC," *m22556*, Geneva, Switzerland, Nov. 2011.
- [14] K. McCann, B. Bross, S.-i. Sekiguchi, W.-J. Han, "HM4: High Efficiency Video Coding (HEVC) Test Model 4 Encoder Description," *JCTVC-F802*, Torino, Italy, July. 2011.
- [15] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Doc. JVT-AB203: Draft reference software for SVC," MPEG / ITU-T, Tech. Rep., Jul. 2008.
- [16] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *document VCEG-M33 of ITU-T Video Coding Experts Group (VCEG)*, Apr. 2001.